Deep Science Publishing
https://doi.org/10.70593/978-81-988918-1-5_6

Chapter 6: Artificial intelligence governance in action: Balancing innovation, transparency, and regulatory compliance

6.1. Introduction

Governance has traditionally been defined as the act of directing, controlling, and managing the operations of individuals and organizations to establish and achieve desired, predetermined objectives. As societies have grown more complex, interdependent, and diverse, governance has extended beyond corporate governance to that of groups and networks. In today's world, distributed governance allows a variety of people and groups to participate in decision-making processes. Although wealth, power, and resources are not evenly distributed in society, people can assert special interests and collective ability to pursue specific social, economic, and political goals. Distributed governance frameworks create the context for cooperation and collaboration to solve local, national, and global challenges, reinforce rules, and nurture relationships (Borgesius, 2018; Cath, 2018; European Commission, 2019).

AI has the potential to accelerate innovation and add significant value to the economy while also posing substantial risks to individuals and society. The economic, social, and political ecosystem around the development of AI is diverse and complex, involving a multitude of stakeholders and dimensions, including research, technology, economy, business, regulation, law, politics, culture, and civil society. How AI systems are designed, developed, used, and managed will determine whether their impact is positive or negative. A cooperative and coordinated multi-stakeholder approach will create the conditions necessary to maximize the opportunities associated with AI systems and protect against associated risks. The promises and pitfalls of AI need to be governed in such a way as to balance incentives for responsible innovation and regulatory compliance. AI needs to be governed by rules for the curious and constraints for the reckless. Without a governance foundation on which all of the actors can effectively build and innovate, maximizing positive societal impact will remain tenuous (Gunning & Aha, 2019; Jobin et al., 2019).

6.2. The Importance of AI Governance

AI governance is hinged on the principle that an ever-expanding reliance on AI systems mandates increasing responsibility for our society in both building AI systems and using them. Governance is important to ensure that AI systems are built and used in ways that align with the goals of our society. AI governance thus states that we are only justified in putting AI systems to use in society in a way that takes into account collective moral considerations, weighing the harms and benefits of putting particular AI systems to use, and that we are only justified in continuing to use them when they are seen to be both safe and effective. Increasing reliance on the use of AI systems that directly impact humans implies an increased need for regulation and accountability, and poses challenges for the implementation of those requirements relative to the governance of other technologies. This is especially true as the scale, capability, and autonomy of AI systems increases. Governments and corporations are finding it hard to balance the humanitarian, ethical, economic, and legal perspectives behind developing and integrating AI technology into society, while risking loss of privacy, social control, employment, or the inability to hold someone accountable for decisions made by an AI system. The current landscape of AI regulation and governance is a mixture of ethical guidelines, policy frameworks, and laws. However, with current societal power dynamics of concentration, bureaucracies are at a disadvantage relative to corporations. The latter are able to spend enormous resources on lobbying efforts to block the promulgation of regulations that might put them at a disadvantage relative to competitors. Corporate actors also have the ability to internally influence the actions of the people who write the regulations around the use of AI in society.

6.3. Current Landscape of AI Regulations

The evolving understanding of the societal risks posed by AI technologies and systems has led to a variety of attempts to conceptualize and, to some extent formalize how to regulate AI. Among the most prominent, spanning multiple geographies, are the Principles on AI, the Hiroshima AI Process, the Coordinated Plan for Artificial Intelligence, and the AI Act.

The Principles reflect a shared recognition of the economic and social impact of AI, and the need to tackle AI’s challenges while improving its opportunities and benefits. The

Framework provides for a growth-oriented economic approach, promoting pro-innovation requirements, as well as offering suggestions on specific policy implementation.

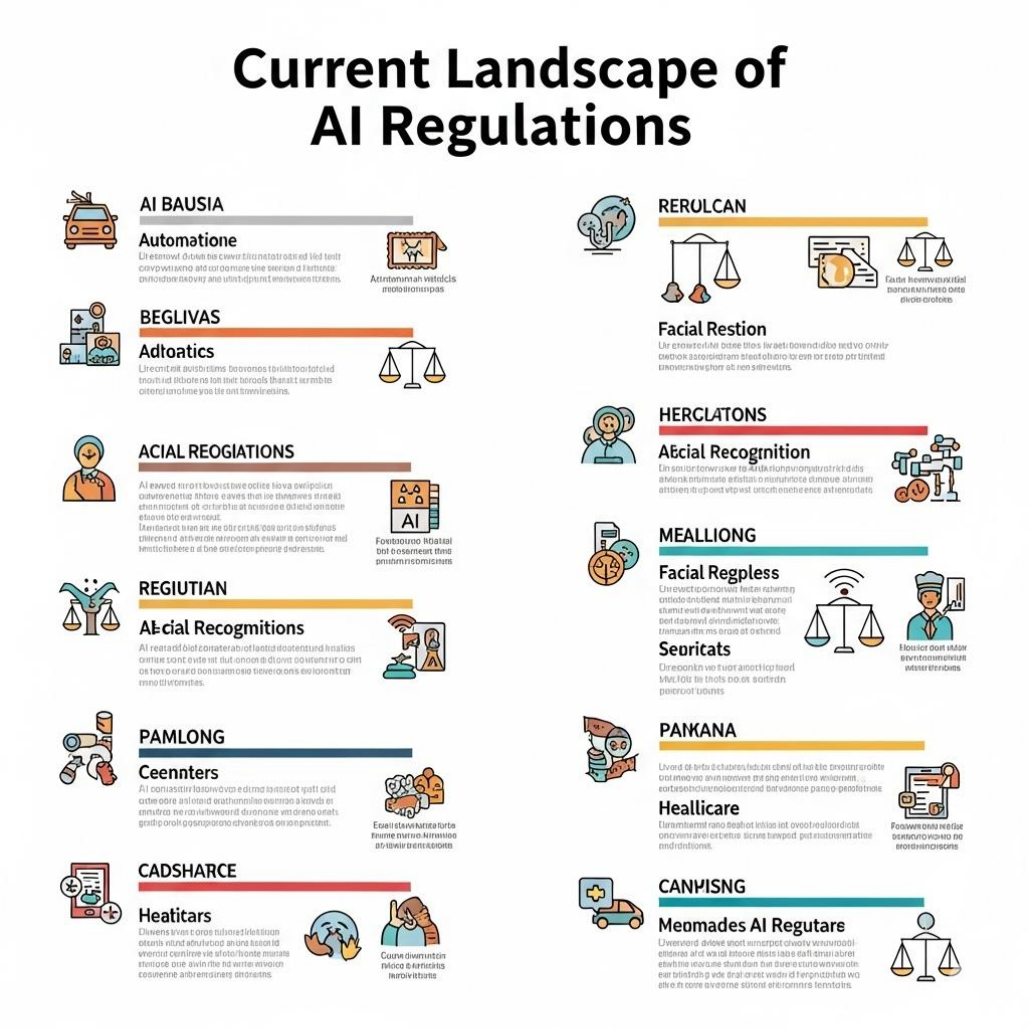

Fig 1 : Current Landscape of AI Regulations

The Summit provides a more simplistic Quid Pro Quo framework of Industry and State. As of now, only the United Kingdom has consolidated its regulatory proposal into the draft Artificial Intelligence Regulation Bill, officially presented to the Parliament in March 2023. Overall, the HAP recognizes a limited number of areas where governments will need to implement governance measures, and these measures are primarily of the Hard Law type that would follow the traditional ex-ante or ex-post models of regulation policymakers are accustomed to use.

The AI Act has its roots in the White Paper on AI, published in February 2020. The White Paper was accompanied by a Public Consultation from February to May 2020. Following the responses to this feedback mechanism, the AI Act sets certain "horizontal" requirements for all AI Act systems, to be implemented by various AI operators, such as "general-purpose AI developers" and "users" who throughout the lifecycle of an AI Act System will be responsible in different ways for its compliance with the requirements set by the AI Act.

6.3.1. Global Regulatory Frameworks

The past years have seen a rapid increase in global efforts to establish regulatory frameworks for AI systems, from unilateral initiatives to bilateral and multilateral partnerships. The European Union has proposed the AI Act, the United States is working to put in place a risk-based regulatory framework for high-risk AI systems, while other countries such as Australia, Canada, China, Japan, Singapore, South Korea, Switzerland, and the UK have developed national strategies and guidelines. Recently, headline events have underscored the ongoing international dialogue between like-minded countries focused on an open and secure AI technology ecosystem.

International bodies are also increasingly taking on a role in AI governance. In December 2022, the United Nations Educational, Scientific, and Cultural Organization adopted the first global standard for AI ethics, based on the unique multilateral experience. The Council of Europe is advancing the work of the Ad hoc Committee on Artificial Intelligence, assisting member states and other stakeholders in the implementation of the policy recommendations and the development of a legally binding instrument for AI. General Assembly Resolutions have also called for further work. The International Telecommunication Union is also active, with the establishment of global AI standardization focus groups in areas such as internet and AI to accelerate digital inclusion.

6.3.2. Regional Variations in AI Policy

Differences in approaches to AI regulation are quite pronounced, given differences between economies and cultural expectations. For example, while the EU clearly places more value on ethics and human rights-based approaches, the US states that economic competition and innovation are primary goals for AI policies. The US outlines the role of the federal government as a 'supporting actor' in AI development. It requests recommendations for specific large-scale funding to support scientific discovery, innovation, and economic competitiveness in priority areas, such as microelectronics, energy, bio-manufacturing, and climate science. The EU seems to prioritize socio-economic policies over industrial policy.

The fact that certain élites in the US—both in corporations and politics—see themselves in competition with the Chinese state and AI sector as the protagonists of a self-fulfilling prophecy of a coming "technological cold war" may confound the perception of many outside the US that use these approaches as a methodology to delegitimize the regulatory initiatives in Europe. They see it as a way of squeezing out European competition. Simultaneously, however, in the US since the start of the new millennium, the growth of the Silicon Valley tech economy has fueled a broader neoliberal and libertarian ideological vision of private actors, not only acting as the key innovation drivers, but also as natural stewards of the public interest, determining the rules for themselves.

6.4. Key Principles of Effective AI Governance

AI Governance defines the broader expectations for the development, deployment, and operation of trustworthy AI systems, while also ensuring that AI innovations are harnessed for the collective benefit of society. Good governance does not interfere with beneficial innovation; rather, it incentivizes companies to prioritize safety, quality, and ethics when designing AI products while also ensuring that there are clear penalties for wrongdoing that affect the welfare of society more broadly. The most successful frameworks balance innovation and risk mitigation, delineating a space of allowed actions while establishing true north principles that minimize the potential for unintended consequences. One of the main challenges in governing AI is determining what constitutes appropriate behavior for different actors. For example, should the expectations for tech developers differ from those for end users? What weight should we give to profit-seeking organizations versus non-profits? Inorganic organizations such corporations generally have clearer accountability, as there are designated leaders who are stakeholders in the consequences of actions taken on behalf of the company. Yet these stakeholders may still seek to distort outcomes to favor short-term financial gains over societal values. The organizations that tend to align well with societal values are non-profits and charities, but they have less incentive to abide by human-centric AI design principles. Individuals, specifically end-users, are often the most affected by AI applications, and are also best positioned to provide input on expected behavior.

6.4.1. Transparency

The principles of transparency, accountability, fairness, and privacy protection are core values that must guide an effective AI governance framework into practice. Without explicit consideration of governance principles, machines can make decisions that contradict our understanding of good decision-making, and business, government, and society may suffer significant negative utility. Our understanding of good decision-making balances multiple conflicting ethical values. While AI has the potential to vastly increase efficiency, conserving resources to pursue altruistic purposes, it can also make careless decisions without regard for social benefit, like the business that discards discarded sheet metal that could be recycled to reduce waste. Ensuring that these transparency principles are adhered to produces technological systems that respect all community values, especially in high-risk domains such as health care, finance, and law enforcement.

Transparency means that AI must be visible to regulators and overseers. There are a number of good reasons for government to mandate transparency requirements. Transparency exposes the workings of the AI system to scrutiny by third parties, which provides additional opportunities for the identification of harmful biases. Transparency provides government feedback on the performance and consequences of AI in the wild, which can inform the legislative process. Transparency can provide information to potential victims of the AI system. If government cannot comply with transparency mandates, it can provide citizens assurance regarding policies, practices, and technologies that mitigate unfortunate consequences of a lack of transparency. Many of the use cases for AI regulation are domains that are highly sensitive to particular social consequences. Therefore, government must take these consequences seriously. Transparency can facilitate the assessment of unintended consequences, so that after-the-risk analysis and remediation can be accomplished more effectively.

6.4.2. Accountability

Accountability in the AI governance context furthers the purpose of systems built on or for the public good and who has the responsibility to act on the findings of audits and assessments. While transparency and accountability are closely linked, transparency alone cannot assure accountability. Governments and the private sector must be held responsible for any harm caused by these systems. While systems have been audited, through the publication of information or other measures, governments and the private sector must ensure they are held accountable for the implications, outcomes, and consequences of accurate and inaccurate systems, both intended and unintended. In addition, accountability should require that historically ostracized communities must be actively and intentionally engaged in the building, use, and monitoring of systems that impact their lives.

Given federal responsibility toward the Constitution and federal civil rights laws, the federal government must take the lead to ensure that all systems are accountable, and that effective monitoring and assessment occur. But in addition to ensuring monitoring of state and local agency use, the federal government must assure that the standards being used are based on best practices from the field in the same way that the federal government assures that the policies on police use of force are based on the best knowledge to protect lives. New forms of private public partnerships are needed to ensure compliance with civil rights, civil liberties, and human rights laws. Consent decrees with public reporting requirements following the pattern of those governing police oversight and use of force are one way to ensure grievance protections and compliance.

6.4.3. Fairness

What we mean by fairness can depend on many factors including the context, jurisdiction, and individuals' personal beliefs and upbringings. In the context of AI that may affect by way of decisions or recommendations or outputs in some way sensitive application areas such as credit, criminal justice, education, employment, housing, health, insurance, and welfare, most people will expect that fairness means some form of nondiscrimination so that people are not treated differently due to being in such sensitive demographic groups as race, ethnicity, religion, sex, gender, national origin, age, or disability. In many jurisdictions, laws recognize a subset of these groups for proscribing discrimination in various areas such as employment, housing, and lending.

A wide variety of algorithmic techniques have been put forth for attempting to ensure that AI systems are fair in this nondiscriminatory sense. However, whether or not a given nondiscriminatory approach is appropriate for a given AI system obviously depends on the context, as well as the particular notion of fairness and bias in that context. For example, a disproportionate number of recommendations for arrests for a particular group are not necessarily unfair if the crime is prevalent in that group, and a disproportionate number of negative recommendations for loan applications for a particular group are not necessarily unfair if that group has a high credit default rate for loans. Thus, it is essential that when designing AI systems in sensitive areas, sufficient care be taken to choose the appropriate notion of fairness for the particular application.

6.4.4. Privacy Protection

The growing use of AI is introducing new challenges to personal privacy, and lawmakers, organizations, and citizens are scrambling to find the right solutions. New AI technologies are being adopted by organizations in a variety of fields and for multiple use-cases. From embedding image generators in productivity tools to using text-based tools for summarizing and writing tasks, examples abound. Meanwhile, portions of the public are enthusiastically uploading personal data into these systems. This provides a trove of user-uploaded, but sensitive or confidential, high-quality real-world data for training data scrapers and potentially for AI developers. Data uploaded can also heighten concerns about bias and toxicity in Generative AI systems, as the systems are trained on different kinds of user-uploaded data and as users interact with the systems.

Countries outside have concerns that privacy regulation could hamper AI deployment and economic competitiveness. And, in the EU, concepts of sensitive data, such as the special status of genetic, biometrics, health, racial or ethnic origin, are being integrated into AI regulatory policy. Currently, the discussion of data protection in the AI Act is incorporated primarily in two categories of requirements: The first is the transparency obligations on AI providers and users. The second procedure and processing principles would give individuals more power concerning their private information, such as their right to be informed that their data will be used to train an AI system, and concerning their contact with AI systems.

6.5. Challenges in AI Governance

In this section, we elaborate on some of the challenges in governing AI systems, including technological complexity, rapid innovation cycles, and stakeholder engagement. These challenges raise the cost of AI governance, and hence we may observe less than socially optimal amounts of AI governance at the organizational level, potentially necessitating government intervention to make sure that the levels of AI governance at the society are sufficient.

Technological Complexity

A fundamental problem of AI governance is that AI systems are technically complex and complicated – they are often referred to as Black Boxes, due to the presence of complex models such as those based on deep learning and also due to the fact that developing and deploying AI systems involves a number of components that are themselves technically complex and complicated. This makes it difficult to govern AI systems adequately. In particular, even if non-expert managers and users of AI systems are keen to govern them properly and are aware of the channels through which an AI system can cause harm, the actual processes leading to harm caused by such AI systems are so complicated that it may be beyond the capacity of non-expert managers and users to pinpoint bias, transparency, explainability, security, privacy, or human oversight issues with the AI system. Further, even if some managers and users of AI systems develop the skills to recognize the presence of the aforementioned issues, doing so for every instance of AI deployment will be burdensome, and hence they may not be extra vigilant and may not know which instances of AI should be governed carefully. For practical purposes, we may say that a suitable indicator should be developed that can be trusted by non-expert managers to ensure that whenever the indicator triggers, the AI system calls for expert intervention, and if the indicator does not trigger, it means that expert intervention is not required. Such an indicator is critically important as organizations start using AI more widely in everyday operations.

6.5.1. Technological Complexity

Technological development is often a very complex matter; it typically builds up on decades of previous work and relies on scientific fields that are only vaguely acquainted with each other. The development cycles for various sectors - even industrial ones - can take some years or even decades, during which technologies are being developed that society is not yet aware of but that promise massive changes in some time. There is a broad breadth of involved stakeholders, ranging from researchers in universities and private firms, who do most core AI development to users of AI services, to governments and NGOs who call for certain use cases of AI technologies, who may fund AI work or restrict it, or who are concerned with the ethical grounding of this research and its application. In this plethora of stakeholders, there are also some actors who may not be well informed or responsible, and whose concerns or guidelines may not fully reflect the larger dynamics of innovation cycles.

Due to the nature of advancing scientific knowledge, some areas of industrial research are highly specialized as they rely on a very specific sub-field of AI technology. In this case, progress will rely more on long-term funding but also on the inclusion of other stakeholders such as climate researchers who specify the requirements, and who actually apply the AI-assisted technology. However, the core developers also need to be aware of the domain-specific needs in order to be able to provide reliable products. For other areas of AI deployment, including those affected by regard regarding safety and fairness of algorithms and model results, it will be challenging for core developers of the involved AI techniques to provide generally guaranteed products without exact knowledge of the application domains.

6.5.2. Rapid Innovation Cycles

While a few simple iterations can happen rapidly through open-source ecosystems, many other technological pathways to AI capabilities are complex, often involving many iteration cycles across multiple technology vectors, including the training datasets, architectures, training algorithms, and output evaluation criteria. Especially for multi-modal systems with ever-changing requirements from business and user usage, multi-dimensional technological development is to be expected. So while certain classes of capabilities may appear superficially to progress by large leaps in capability due to rapid cycles, it is more likely that the underlying breakthroughs required would be surprise innovations that happen surprisingly fast, while more incremental innovations may take longer.

Importantly, just as individual companies have different velocity profiles for various AI products, so is there a diverse ecosystem of companies with various capabilities and interest in diverse processes. Just as some sectors have seen longer cycles of investment, deployment, and returns, AI has both rapid cycles and much longer cycles, each with opportunities for governments to engage. The implication of this point is that there will be periods of intense interest from both startups and established players innovating in the space of AI concurrency for social, budgetary, and other concerns, and it would benefit governments during these cycles to engage with inquire directly into the incentives and needs of these private sector actors to maximize benefits to society.

6.5.3. Stakeholder Engagement

Governments have extensive experience with public engagement allowing them to pursue a wide range of goals, from informing citizens about policies to involving them in decision-making and implementation. Goal-setting and process design are the keys to efficient stakeholder engagement. In the example of warning labels, policy-makers could simply have mandated their use across the board. However, particularly in the EU, such a blunt tool would have been used much less. By opening the label design process to public input, policy-makers could harness public interest in the labels to foster awareness of and interest in the information behind the labels. The labels could then be made a more effective tool for informing citizen consumers.

In the case of AI, some stakeholder groups are more likely to be affected on the basis of their personal characteristics than others, while in some cases patterns can be observed at a more general level, such as for countries using automated systems to monitor the activities of classes of people. Additionally, the severity of harm could also differentiate who should be explicitly consulted for input. Sadly, it is often the most vulnerable who are affected by the potential malicious uses of AI, or whose society lacks the capacity to engage with a consultative or collaborative process to inform budgets and priorities. Policy-makers should ensure that input is derived not only from a focus group, but also from those who might be able to generate the desired outputs.

6.6. Balancing Innovation and Compliance

Innovation represents a competitive advantage, but can also violate rules. There is an inherent tension between industry and governmental regulators, where companies push for less regulation, while the government implements more to gain more tax income or other forms of political capital. An AI developer has to be innovative in some areas in order to remain on the technological cutting edge. In this process, existing regulations may be violated, while companies also seek to innovate in other areas in order to cover their revenue and profit goals. Regulatory agencies are left with the challenge of not stifling innovation through too strict regulations, while creating enough guidance for companies to comply with regulations. Companies should focus and prioritize on those regulatory areas that would provide the biggest benefit to society. Focusing return on investment on ethical and compliance governance matters considered critical by stakeholders and regulators alike is the only viable path. Similarly, regulatory agencies should favor a risk-based approach to compliance. This could be achieved through phased compliance requirements, whereby easier, first steps are required for a company to engage with regulators and stakeholders, while gradually moving the company into a more transparent and thus complex compliance posture. It is quite possible that over time, communication between the agency and the company would improve considerably and create benefits to both sides. The tremendous knowledge and innovative process expertise of companies can provide regulatory agencies with information on how companies employ their expertise, what critical paths they follow, and in which ways those paths may get substantially derailed. Companies may also be encouraged to provide recommendations on how best to manage the duties of the agency.

6.6.1. Fostering Innovation in a Regulated Environment

AI technologies offer significant benefits to society. They can improve economic productivity, address key challenges such as climate change and disease, and augment human perspective and problem-solving capabilities. The objective is to become the global hub for cutting-edge, trustworthy AI – meaning AI that is ethical, safe, secure, and respects human rights. The goal of any AI policy should be to ensure leading positions in developing and deploying cutting-edge AI technology; while safeguarding the public good and building trust with its citizens. Governments can play a role to spur development and use of beneficial AI technology through research funding and through ensuring a favorable economic environment. Starting in their first years of operation, the current U.S. Administration and European Commission have both prioritized investment in AI and machine learning in their respective R&D budgets. Europe and the United States can further spur beneficial technology adoption with supportive tax codes and smart public procurement.

Fig : Fostering Innovation in a Regulated Environment

However, beyond regulating bad conduct and bad outcomes, there are limits to what governments can do to create innovation. There is no formula for how businesses will create new AI technologies or what form they will take. There is no predictable schedule for how businesses will adopt technologies, or who will be the first movers. On the innovation and commercialization sides of the equation, policymakers need to accept that successful actors will not always be companies that are the largest, or that are American or European. Neither the United States nor Europe can predict how and when businesses will decide to build the leading – and value-creating – companies focused on the innovative commercialization of these technologies.

6.6.2. Case Studies of Successful Compliance

U.S. companies calibrate compliance with the various overlapping constitutional, statutory, administrative regulatory, and common law requirements, determining that compliance in a relaxed manner is sometimes best. For example, investment in new technology has been encouraged by light touch regulation allowing drones to operate in restricted airspace from which they would be practically excluded by rules.

Particularly impressive instances of careful compliance can be found in the reports of successful companies prepared by law firms. These cover cutting-edge health technology, fintech companies, and distance learning undertaken by major educational institutions. In price and quality of new products and services, complying firms' offerings are sometimes superior to those of companies who choose to ignore regulatory requirements. The distinct advantages enjoyed by law-compliant companies (apart from avoiding risks of civil or criminal liability for regulatory infractions) is their use of branding and information disclosure to advertise to would-be customers their business activities in compliance with applicable regulatory regimes, including privacy, cybersecurity, and anti-money laundering regulations that govern technology-driven interventions.

Compliance provides law-abiding marketplace actors, who incur the expenditures necessary to bring their activities within the scope of regulation, with considerable reputational goods. Social media have given a megaphone to complaints about industrial practices, whether dolphin- or child-unfriendly. EcoWarriors used to provide only small statures for spoken complaints, until recent years when the exposures available online have raised awareness of, and condemned, junk research as hate speech.

6.7. Best Practices for AI Governance

AI governance is a nuanced and complex endeavor that will balance unique innovation challenges while ensuring compliance with unified ethical principles set out in organizational AI policies. Defining common sense organizational-wide decision-making tools, processes, and controls, while remaining flexible enough to accommodate distinct business unit models is the best practice approach to enable dynamic AI use. Companies that take this approach will be able to establish bottom-up sandbox AI innovation use cases for limitless business unit models while staying compliant within defined policy guardrails. For AI to offer the greatest benefits, organizations should establish and monitor a defined set of flexible and dynamic governance frameworks that provide clear guidelines and shared understanding of the organization's expectations. Communication is key and should begin at the onset of any AI development. Clear AI ethical principles developed at the organizational-level will help guide the establishment of bottom-up governance framework definitions by business units. These principles may touch on issues related to privacy, security, consent, quality, and explainability and should be clearly laid out and communicated within the organization. These ethical frameworks are the backbone to greater efficiency and will help guide joint development processes to validate alignment to be sure AI tools are devoid of bias and fulfill ethical requirements within the defined policies. Ultimately, these policies develop trust in AI use within the organization. Trust fosters successful collaborative partnerships for AI model training, implementation, and continuous improvement of AI.

6.7.1. Establishing Governance Frameworks

Artificial Intelligence (AI) is a set of technologies that enable machines to perceive the world, reason, learn, and take action. AI has great potential to improve the human condition. Yet, AI also introduces novel challenges and risks that must be managed in an ethical and trustworthy way. These risks include but are not limited to the perpetuation of societal biases and discrimination, the undermining of fairness and accountability in decision making, and the jeopardizing of privacy and security. Many of the AI systems deployed today in sensitive domains such as criminal justice, education, or healthcare are routed in complex and opaque algorithms. These systems may not only produce inaccurate outcomes that could have life-altering consequences. Misjudgments can also go unchecked, due to a lack of transparency, oversight, and redress for affected populations. Recently introduced or proposed policies, frameworks, and principles to realize trustworthy AI seek to foster a development of AI systems that minimizes such risks and harms.

Policies and guidelines often alone are not sufficient to achieve this goal. They should be paired with concrete measures that organizations adopt to try to realize the principles set out in the documents. In particular, organizations should create implementable ethical principles on designing and deploying AI systems and continuously assess the impact of those systems as they get used in the real world. They should perform stakeholder engagement and design needs assessments prior to deployment of ethical standards in projects. In addition, organizations should conduct usability tests to evaluate whether the ethical guidelines are being followed by the AI systems used in production. These enhanced ethical guidelines are a necessary part of an organization’s AI governance framework. These documents specify how organizations operationalize their ethical principles in practice.

6.7.2. Implementing Ethical Guidelines

To ensure that AIs are developed and operated responsibly, organizations can use ethical guidelines around the developed or operated AI systems. These AI ethics guidelines aim to create a framework to structure organizations’ decision-making with respect to these systems. Actual in-depth examination of real-world issues and challenges that affect accountable behavior, potential unintended consequences and regular revision of AI governance systems and associated mechanisms are instrumental in achieving and maintaining public trust in AI systems. Creating an AI ethics guideline involves collecting information about the potential risks that various people face while AI systems would be developed. Each of these guidelines describes one or a specific set of principles that the organization will adhere to while taking decisions around the AI systems. As a next step, it is necessary to collect further detailed information about these principles. Potential requirements can be derived from giving specific meanings to the principles that the organization adheres to. It is further necessary to create a tailored implementation plan based on the identified requirements. In the plan, it is necessary to specify further concrete choices that can be made concerning the identified AI guidelines to provide actionable and realizable next steps we would make when refining those guidelines. Further, some guidelines might have more ramifications than others. Consequently, the governance implementation plan might vary depending on the predefined AI principles. Ethics guidelines may also include a commitment to human oversight and accountability throughout the AI lifecycle, support for stakeholder participation, and consideration of the potential impact on the well-being of all people, not just end users.

6.7.3. Regular Audits and Assessments

Governance frameworks should not be static and must adapt to the rapid pace of AI development. Therefore, it is important to continuously evaluate the effectiveness of governance measures. Regular audits and assessments of AI systems' impact, effectiveness, and performance are essential for transparency and accountability. Companies must assess systems and models continuously to identify and mitigate potential negative impacts. Conducting impact assessments over the AI lifecycle enables organizations to make informed decisions about AI systems. Before, during, and after deploying AI systems, organizations should assess the impact on different people, groups, and environments. Furthermore, as AI systems evolve, organizations should engage with impacted stakeholders to understand if there have been, or could be, any unintended consequences, positive or negative.

Model cards should summarize the intended use, evaluation data, and metrics, as well as be developed as part of a collaborative process with key involved stakeholders. AI developers and deployers should establish disclosure procedures for transparent systems, especially in high-impact use cases. Providing humans with insight into the decision-making of AI models including decision justifications and explanations can help to identify problems with deployed systems. Organizations should also share relevant disclosures in public and with affected stakeholders, drawing from standards where available. The importance of performative action of organizations regarding AI disclosure is amplified in high-impact use cases. Third-party reviews and evaluations of high-impact systems through audits should be made possible and encouraged, compliant with confidentiality, privacy, and security requirements, for accurate and trustworthy assessments of such decisions. Organizations should enable third-party access to information that allows for independent auditing of disclosed products, including data, documentation, and information, while balancing the trade-offs with respect to confidentiality, privacy, and security.

6.8. The Role of Stakeholders in AI Governance

AI governance is a complex environment, requiring cooperation and engagement among a variety of stakeholders, including government, private sector, civil society, and the technical community. These groups bring unique insights on how to harness the value, and surmount the challenges of AI, and have differing responsibilities in ensuring that responsible practices are utilized throughout the lifecycle of AI technology. The varied nature of the stakeholders, and their capabilities and responsibilities for governing AI change in different phases of technology maturity as well as by market character, and the interplay and cooperation amongst stakeholders needs to be emphasized.

Government and Regulatory Bodies

Major responsibility and accountability for deployment of safe and socially beneficial AI rests with government and regulatory bodies. They have the responsibility for ensuring compliance with the ethical guidelines, values, and societal norms. National policies provide the vision of the kind of society as well as the culture within which technological advancement is to take place. Regulatory bodies translate policy into robust safeguards. Governments are also responsible for enabling innovation, nurturing alternative models of development, and making public investments at the national and global levels to address climate change, disaster risk reduction, health, and education.

Private Sector Contributions

Business houses are responsible for implementing ethical guidelines, norms, and values on the ground. They define, design, and build the AI technologies and solutions that deliver business value. Employees at all levels in the organization should be the torchbearers of ethical values. The whole ecosystem, including vendors and suppliers, must be taken along. There should be investment in building safety measures and accountability, keeping in mind the worst-case consequences. Many private businesses are already contributing to the field of AI ethics. Co-creation, research collaboration, and sharing of findings can promote the process of realization of safer and socially beneficial AI.

Civil Society Engagement

Civil society has the responsibility of monitoring and speaking truth to power. They mobilize protest to keep the focus on those who fall between the cracks of a rapidly changing technology landscape. By standing ready with the people who lose out, and advocate the cause of inclusion, they build trustworthiness and contribute to the creation of stronger institutions capable of keeping the flame of democracy and economic parity alive.

6.8.1. Government and Regulatory Bodies

Governments are also among the first to call for AI governance. In part, this is motivated by the implication of significant negative externalities, costs incurred by citizens who are not the users of a system, as well as the effects of technology monopolies who profit at the cost of taxes and other public goods that require a thriving society. A concentration of power among a small number of technology companies can lead to their lobbying interests swaying local and national policies and reducing the accountability of decision-makers. For those businesses that do transact with public institutions, it leads to a potential ‘capture’ of these contracts affecting the accountability standards of their performance. Governments often want to address concerns of how these economies may impact the distribution of wealth whose effects are beyond the reach of any corporate social responsibility initiatives. How such concentration impacts national security also becomes a concern. In addition to the risks of surveillance, both externally and internally, the adoption of biased systems can have severe consequences in terms of how international relations are conducted or how certain ages, communities, or other demographic characteristics are branded within and towards the rest of the world. As the world prepares for the next pandemic, many governments are beginning to stockpile protectors against mass footage, including drones, anticipating their use for monitoring society. The close interaction between the private and public domain makes the distinction between the two less sharp.

The challenge for governments is not so much due to the fast pace of technology development but to the cross-border and seamless nature of investigations, data harvesting, and other forms of exploitation that span multiple jurisdictions. At a local level, concerns focus on the information asymmetries between the citizens, affected by any decision, and the decision-makers; at an international level on how to cooperate. If individual states rush to regulate the development and use of certain algorithms for fear of failure or cheating at the next Olympic Games, and thereby stifle creativity, all would be worse off.

6.8.2. Private Sector Contributions

The AI sector is largely driven by private investment, particularly from companies building large language and other AI models. These companies have played a key part in the rapid development and release of new AI capabilities, and they also have great potential to work with governments to elaborate use cases with high societal impact, to develop specific capabilities, and to scale them on their cloud infrastructure and global footprint. AI companies have expertise in large-scale training of AI models, building safe interaction between tech and end users, human feedback for the AI learning process, and deploying and scaling AI services globally. All these roles are crucial to rapidly address governmental AI use cases. Academic researchers have had a critical role in the past decade in pushing the envelope of public knowledge on specific AI technologies. Research enabled by these advances has given rise to open-source reusable models and provided a fertile ground for subsequent commercialization. No privacy nor safety

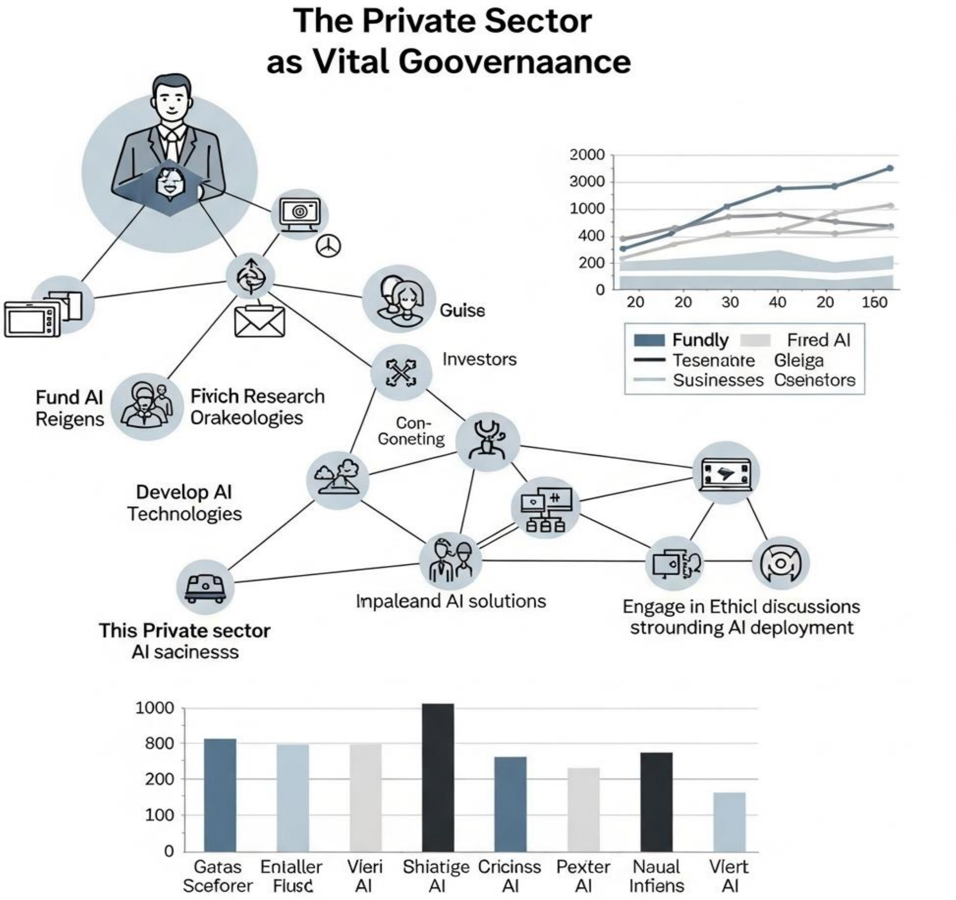

Fig : The Private sector as Vital Governance assessments of AI models have been available, and little defender capacity has been developed to push those companies in the right direction, so private sector social responsibility should compensate for that gap and develop processes to allow for greater transparency and trustworthiness. This responsibility should both support self-governance and organize dialogue with public regulatory bodies, being fully aware of competent areas of such interactions.

6.8.3. Civil Society Engagement

The concept of civil society is multifaceted and has been interpreted in various ways throughout history. For the purposes of this work, we understand civil society to be the private arena outside of economic or state structures, where individuals act together to form organizations, promoting shared interests and aspirations, purposively engaged in processes to positively influence social life. Based on this definition, we consider civil society as consisting of citizens acting individually or collectively, through non-market and non-governmental organizations to influence and/or contribute to the positive advancement of part or all of society's interests.

Many of the subjects of human rights, ethical principles or shared social values are embedded within organizations in civil society, which are always in a close relationship with individuals. In this way, civil society acts as intermediaries between individuals and political authorities. Furthermore, civil society representatives also bear a tremendous responsibility to inform, educate, and heighten the consciousness of their member organizations, in order to identify shared real-life concerns relative to AI technologies and to share common AI principle recommendations that address specific issues. These groups' analyses would allow appropriate and shared demands for action or regulation to be presented to regulatory authorities and public policy makers.

In short, functioning as both sides of the same coin, civil society representatives, by presenting valid demands based on citizens' experiences, play the crucial role in AI Governance by providing authorities with the information they require to develop and implement appropriate policies for the social good, while at the same time political authorities have the responsibility to listen, dialogue, and take action relative to civil society inputs and demands.

6.9. Future Trends in AI Governance

AI governance foundations have been laid through the legal frameworks and guidance issued to date; furthermore, its practical deployment has commenced with the implementation of guidelines, the creation of registries, risk assessments, audits, and expert reviews, as well as assessments by boards of ethical AI use and alignment with values. The future of AI governance will rely heavily on advancements in the technologies being governed, such as the increased sophistication of AI systems, algorithmic transparency techniques, advances towards artificial general intelligence, or new classes of technology like brain-machine interfaces, synthetic biology, or the metaverse. It will also rely materially on how regulators and society as a whole choose to evolve their approaches to governance, in either a lighter touch or a more commanding way, and what trade-offs this entails, both in regard to how such choices impact the types of innovation that are expected, as well as what types of societal impact are being remediated or prevented. Above all though, the resiliency of the proposed solutions, legal frameworks, and guidance requires societal consensus and reconciliation of perspective across the silos within which we tend to think about life and therefore legality and the regulation of technology. Before we can work toward International Law and treaties that spurn a whole of society approach to AI governance, we need to work on harmony and facilitating joint work at the country level.

6.9.1. Emerging Technologies and Their Impact

The role of AI governance has gained momentum and will continue to evolve with the arrival of emerging technologies such as generative AI, quantum computing, and synthetic biology. With the rise of the Internet of Things, humanity is more connected than ever before – and with society drastically shifting how they work, play, and live, it is imperative that the AI governance response develops at an equal and rapid pace. On the other hand, historical intelligence governance has depended so much on the availability of technologies and the understanding of impact, such as the reaction to nuclear weapons or the lack of societal interest in synthetic biology after its arrival and subsequent quiet period. In the AI context, we expect there to be more societal awareness on AI from the societal impacts in terms of disinformation, labor market changes, and other developments that have impacted industries.

In light of new developments in generative AI, governments around the world have fueled the discussion on rapid investment, development, and research into these, while also urging technologists and corporations to ensure that they maintain safety and security. Thus far, we have seen how tech companies have aligned to roll out protective features in terms of algorithms in image generation, warning signals and detection criteria in terms of detection of deep fakes, and detection and protection in their large language models. While these features seek to ensure that this technology can be used in a trustable context, there are still considerations, such as the prevention of cybersecurity attacks using generative AI, and the increasing volume, sophistication, and scale of these tools and services, which may incubate economic volatility and espionage.

6.9.2. Evolving Regulatory Approaches

As generative AI becomes increasingly integrated into our social systems, regulatory frameworks will need to evolve beyond current approaches that focus on self-regulatory guidelines and upticks in managing risk. Regulatory agencies will need to transform how they work to meet the ongoing challenge of ensuring public trust, safety, and accountability while both overseeing and fostering responsible innovation. There are already discussions of approaches that include continued enhancements to risk management standards and dedicated AI regulations, requirements for developers to make their LLMs open source, and calls for the development of AI governance regulators.

While these upcoming approaches look to mitigate risk and their differential impacts on the most vulnerable, look for an increasingly growing call for verification efforts beyond what seems to be acceptable at this stage to apply more of the cybersecurity strategies already in place rather than needing to start anew. There should be a greater focus from AI developers on transparency features regarding what information AI algorithms are provided. While there are efforts underway in various sectors, continue to hold skeptically the assumption by some LLM developers that transparency and verification are not needed as we work together to create AI that best serves our social needs.

6.10. International Collaboration on AI Governance

The cross-border nature of many AI activities creates a need for global norms that can support an appropriate framework for AI development and use. Already, initiatives are being developed, and models drawn from other sectors are starting to emerge, including:

- A recommendation on the Ethics of AI, which calls upon member states to adopt an ethical framework for AI and includes key principles for the Governance of AI.

- Principles on AI, which are included in a declaration signed by many governments and outline the importance of transparency, accountability, and human-centred approaches to AI.

- A framework for the Ethical Use of AI in the Military which outlines possible philosophical, ethical, and legal use considerations for international partnerships.

- Approval of key global partnerships including the Global Partnership on AI, the Coalition for Digital Trade, and the D10 partnership for technology which all invite shared values in their frameworks.

Additionally, a consensus on the “Future of Humanity” where the key risks from increasingly capable AI systems are addressed.

6.10.2. Cross-Border Regulatory Challenges.

Exploring the content that underlies existing initiatives and partnerships begs more questions: how should regulations be approached across borders? Which sectors are best regulated at home or abroad to minimize regulatory burden? Will country- or region-specific requirements make it impossible to consider cross-border UX/UI consistency? Will players in the AI ecosystem be able to support business model viability if revenue sources and support structures are not harmonized globally? What impact will politics have on collaboration?

As these collaborations and international discussions unfold, companies and innovators should be aware of the implications on their business operations and product features. Keeping on top of the risks and support structures being built today will support more informed decisions on the horizon.

6.10.1. Global Initiatives and Agreements

As we witness the dizzying upsurge of generative AI technologies and their applications in many industries, we are simultaneously reminded of the need to develop a comprehensive infrastructure of agreements and institutions to promote the development of artificial intelligence models that balance their many advantages with the real risks they pose to society. Topics to be addressed in this global infrastructure involve, in a non-exhaustive manner, the assurance of AI's technical robustness and safety; the mitigation of any potential negative impact of AI on people's life or that of the planet it affects; the assurance of transparency and responsible use of AI systems; the promotion of shared values and the safeguarding of human rights; and the establishment of accountability mechanisms for the enforcement of these principles. At the same time, the buoyant development of AI technologies also reminds us of their transformative potential to foster economic growth, social progress, and other blessings we aspire towards. Therefore, a healthy collaborative global governance mechanism would also need to recognize the positive potential of AI.

Initiatives to collaborate on the governance of AI are starting to be created at the international level. A good part of these initiatives is building on previous work on governance in the domains of data and digital security. The widespread use of deep learning models based on data and computational power creates a functional necessity for engagement between major actors. All countries not working to build their own capacities are in effect relying on the efforts and outreach from major technology companies. It is likely that investment and other incentives would be created to support the sharing of AI systems, competing to build innovative application systems that further the common good, and the responsible use of large-scale models.

6.10.2. Cross-Border Regulatory Challenges

The global scope of AI innovation, investment, public usage, and economic impact raises a variety of complex challenges for the governance of AI at national, regional, and international levels. Laws and regulatory policy are historically established, informed, and implemented at the local level within countries, and are intended to reflect the unique characteristics, values, and needs of individual nations, communities, and economies. As a technology that is inherently transformative and disruptive, AI can in some cases erode or frustrate the interests of national governments and regulators, as well as community stakeholders, even during the rapidly iterative activity involved in the technology's development and deployment. These effects can involve economic threats related to welfare loss and public lost revenues in the form of taxation; equity threats related to the exacerbation of social inequality; privacy and security threats associated with data misuse or cyberattack; political threats representing the risk of disruptive fake news generation or propaganda dissemination through deepfake technology; and safety threats associated with the unregulated deployment of AI in physical or virtual environments.

While individual nations may attempt to address these challenges through regulatory mechanisms such as input market isolation, trade restrictions or bans, product and service market restrictions or bans, or government subsidies for local producers, the benefits to economic and national security strengthening from collaborative engagement in global trade and technology partnerships may outweigh the associated challenges. The associated deployment or commercialization risks without effective policies for stakeholder engagement of AI, when left unmitigated, may include the lowering of community trust in conventional regulatory policy; the elevation of AI-associated security risks to national levels without international alignment; the misunderstanding of local political objectives without stakeholder consultation; and the loss of human-centric values acknowledged globally. Furthermore, the potential stress on bilateral relations among countries resulting from regional disparities in the expectations of international standards for ethical visible or sensory AI use-policy governance should not be understated.

6.11. Conclusion

In conclusion, as AI has evolved to be incorporated into a greater number of products and applied in new ways, regulatory uncertainty has mirrored this expansion creating the imperative for innovators to build trust by designing and deploying transparent AI solutions accountable to all stakeholders. This is especially true of the healthcare ecosystem, which is characterized by a combination of factors unique to its engagement with AI technologies. These include inequalities in access to care, the stigma associated with seeking healthcare, the intersection of health and health policy with social determinants, the patchwork of federal and state legislation and regulation, the restrictions of privacy rules, the influence of payors on opportunities to innovate and develop AI solutions, the test-and-learn approach to digital health, and the limitations of AI with respect to regulatory compliance.

These characteristics reinforce the need for innovators to address issues of transparency and trust in a manner that is customized to the particular AI application in question. They include determining what transparency should look like in terms of communicating with patients and physicians and providing safeguards against unintended consequences. They should ensure that access to care and privacy are balanced and that bias, fairness, equity, accountability, and explainability all factor into the design and deployment of their innovative AI solutions as joint assets. In assiduously working to achieve these goals, AI innovators will build trust with patients, physicians, payors, and regulators alike, thereby fostering a culture of collaboration and mutual accountability. Such a culture will encourage innovative thinking about how best to design and deploy AI tools to fulfill the promise of better, cheaper, and more equitable health for each individual and population.

References

A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nat. Mach. Intell., vol. 1, pp. 389–399, 2019.

European Commission, Ethics Guidelines for Trustworthy AI, Brussels, 2019.

F. Z. Borgesius, "Discrimination, artificial intelligence, and algorithmic decision-making," Council of Europe Report, 2018.

S. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," Philos. Technol., vol. 31, pp. 689–710, 2018.

D. Gunning and D. Aha, "DARPA’s Explainable Artificial Intelligence (XAI) Program," AI Mag., vol. 40, no. 2, pp. 44–58, 2019.